Module 2: Assignment 1

Yanyang Pei (B3)

Introduction

In food development, conducting sensory tests is an essential step in the product innovation process, allowing developers to gather valuable insights on consumer preferences before launching new products into the market.

Recently, the food industry has developed interest in organic-based sauces, which are seen as healthier alternatives. This study focuses on a newly developed organic sauce line featuring four flavors, each formulated using different fiber types, including potato, tomato, apple and carrot. And they have different concentrations and are processed under different pressures.

The primary aim of this research is to analyze the sensory test results of these sauces, and the hypothesis is that the fiber type has a significant effect on the sensory properties of the product, especially on thickness.

Materials and Methods

Data analyses were done with Python 6.0. The provided data consisted of a table including 3 factors (fiber type, pressure and concentration) and 5 sensory parameters (graininess, thickness, crispiness, melting and slipperiness), with three replicates for each factor.

The data was examined for outliers with a Grubbs's test, indicating 2 outliers in thickness and graininess. Those outliers were removed and replaced by the average of the replicates from the same condition. Outliers are visualized by line plot as well as scatter plots.

Cluster Analysis (CA) and Principal Component Analysis (PCA) was performed on the whole dataset to study the potential relationship inside the database. These are followed by univariate analysis, including Kruskal-Wallis and bar charts to investigate the statistical difference on thickness caused by different factors.

Results

Pre-Import Preprocessing

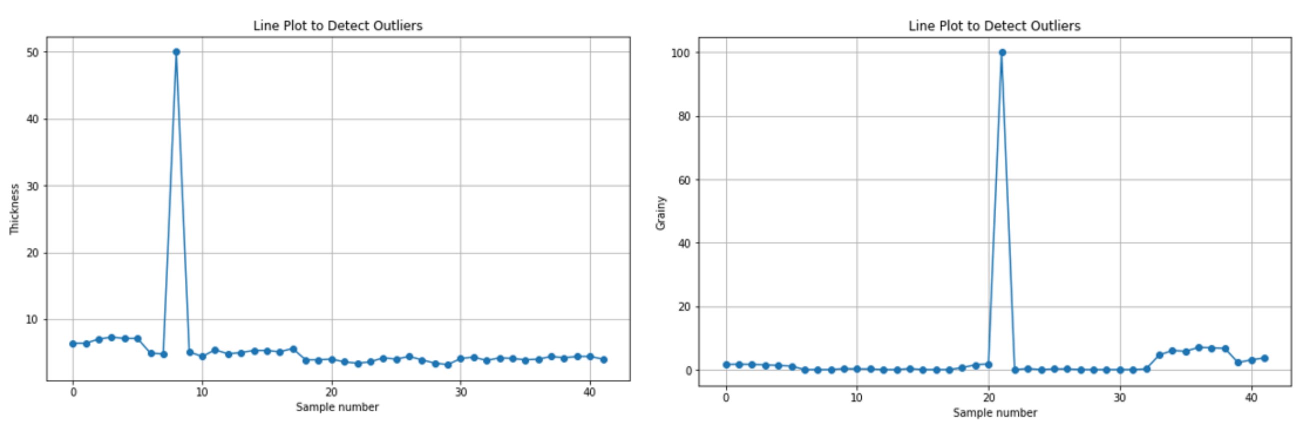
Modification was done to the dataset. A missing row in the table was deleted. There's a blank at melting point of sample 18, and the average of melting point of sample 16 and 17 are filled in. There appears to be outliers in the data, which include the graininess of sample 22 and sample 19, thickness of sample 9, and melting point of sample 34. All of those outliers are deleted and replaced by the mean value from the same treatment. It's also noticeable that replicates 19-21 have large variance in crispiness, but no treatment is performed on it to ensure authenticity. The modification was tested by comparing the result filling with medium, which seems to be reasonable.

Post-Import Preprocessing

To ensure the integrity of our data analysis, Grubb's test was conducted on all dependent variables. The output (shown in figure 1) indicates there's one outliner for thickness and for graininess, which is 50 and 100 respectively, while other dependent variables don't have outliers. Outliers are deleted and replaced by the average of the other replicates. Line plots for dependent variables that show outliers are made and shown as figure 2. And in scatter plots, the outliers are marked, which is shown in figure 3. From which we can easily visualize the position of the outlier in the dataset.

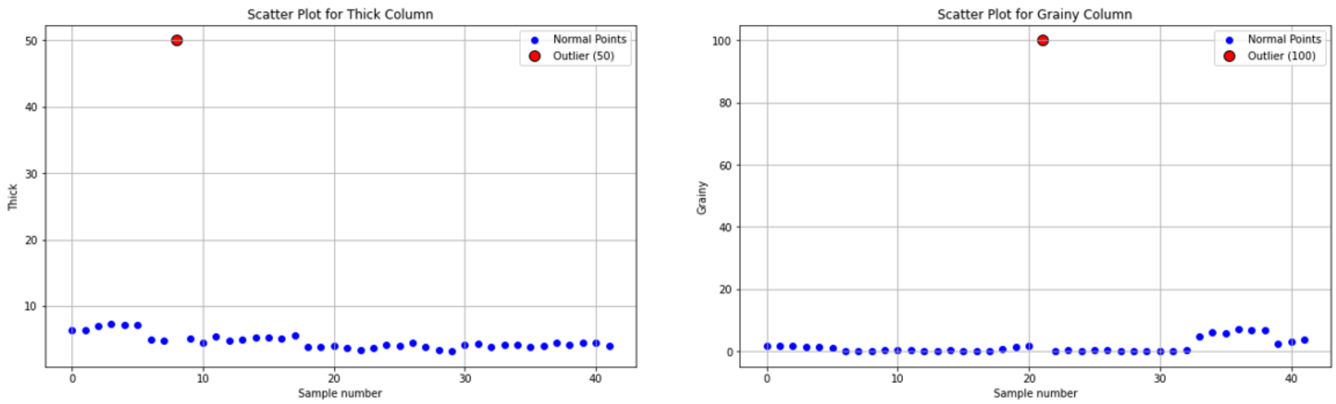
```
Processing column: Thick
Outliers for Thick: [50.0]
Processing column: Grainy
Outliers for Grainy: [100.0]
Processing column: Slippery
Outliers for Slippery: []
Processing column: Crispy
Outliers for Crispy: []
Processing column: Melting
Outliers for Melting: []
```

Figure 1. *Output of Grubb's test for dependent variables*



(A)Thickness (B)Graininess

Figure 2. *Line plot for sample thickness and for sample graininess*



(A)Thickness (B)Graininess

Figure 3. *Scatter plot marking outliner in thickness and in graininess*

The normality of all dependent variables was studied by the Shapiro-Wilks test, and the result is shown in table 1, which indicates all of them are not normally distributed.

Table 1. Shapro-Wilk's test result

Characteristics	P value	Normality
Thickness	$3.4*10^{-4}$	Not normally distributed
Melting	$9.2*10^{-7}$	Not normally distributed
Crispiness	$6.7*10^{-9}$	Not normally distributed
Slipperiness	$2.6*10^{-3}$	Not normally distributed
Graininess	$7.4*10^{-8}$	Not normally distributed

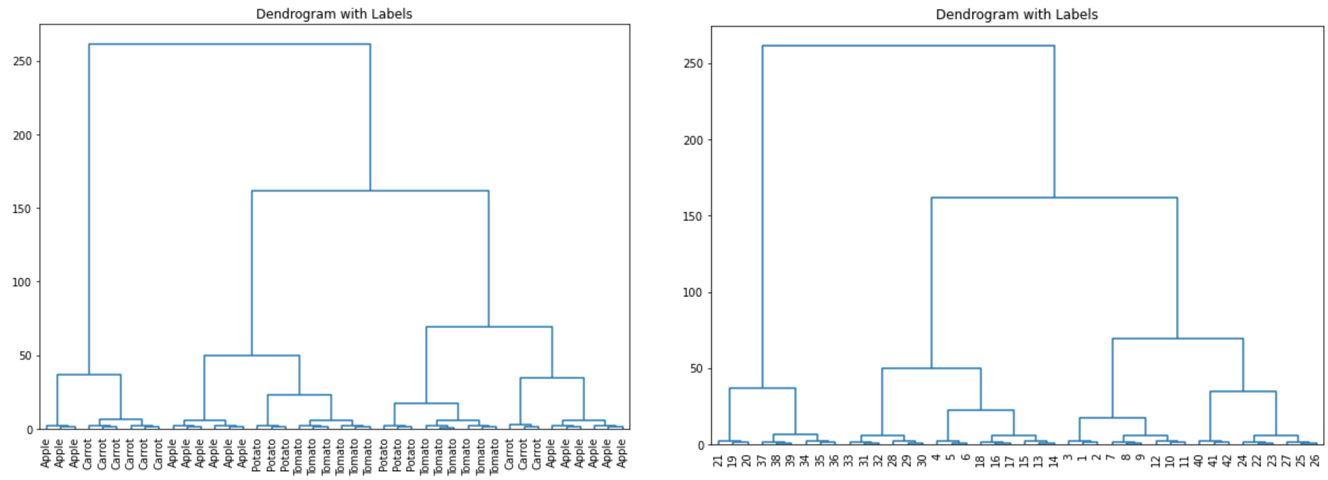


Figure 4. *Dendrogram showing samples relations, where the left one shows the label as fiber type and right one shows label as sample number*

Dendrogram was made to study the similarity of the groups. It can be seen that there's high similarity inside the same fiber types, and medium similarity between different fiber types. Tomato and potato fibers are clustered together, suggesting these two groups are more similar compared to apple and carrot.

MVA on full dataset

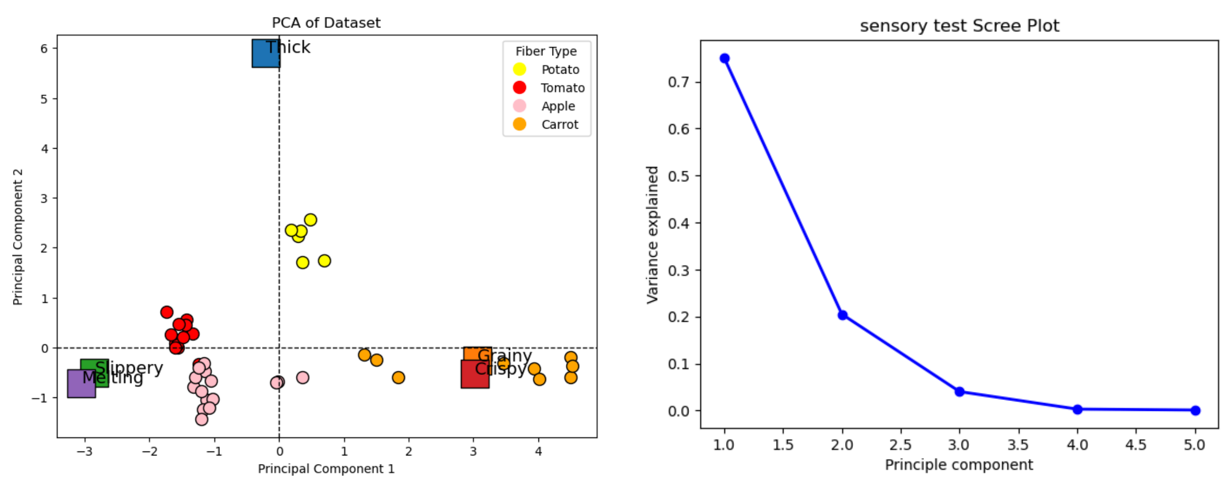


Figure 5. *PCA plot for data set, where round points are from scores matrix while square points represent the loading matrix, and scree plot was shown on the right*

The PCA plot shows the clustering of fiber types and different physical characteristics of the product samples. The PCA reveals distinct patterns among the dependent variables. Notably, the variables slippery and melting are clustered closely together, along with graininess and crispiness, indicating a positive correlation among these attributes. In contrast, thickness exhibits a negative correlation with the other variables.

As for observations, **red points (Tomato)** seem tightly packed, indicating they share similar properties. **Yellow points (Potato)** are mainly concentrated in the top center of the plot (PC1 around 0 and PC2 between 2 and 3), suggesting this group has different characteristics from the others. **Pink points (Apple)** are found towards the lower left (PC1 between -2 and -1 and PC2 between -1 and 0). They are more spread out than the others, indicating a wider range of variation. Some of the points spread to the center of the PCA, indicating they’re less influential. **Orange points (Carrot)** are scattered towards the right side of the plot (PC1 around 3 to 4 and PC2 between 0 and 1), with a moderate spread in the data.

For PC1, we can see big differences between sample melting point, slippery, crispiness, graininess, as well as different fiber types.So it might represent these properties of the samples. While PC2 there's a big difference between thickness and other characteristics, hence can represent if the product is thick or not.

From the scree plot we can see that the first two principal components explain more than 90% of the variance, in which the PC1 explains more than 70%, indicating one variable is making a huge impact on the data.

It also appears that tomato and apple samples are close to slippery, while carrot samples are bound with grainy and more crispy and potato samples are bound with thicker. It’s predicted that those properties are related to the fiber types. This is further supported by the following Kruskal-Wallis statistical study.

In addition, by dividing the fiber type into vegetable group (potato and carrot) and fruit group (apple and tomato), it can be seen that the fruits product are clustered closer to slippery and melting, while some vegetable products are spreading wider towards grainy, crispy and others are close to thick.

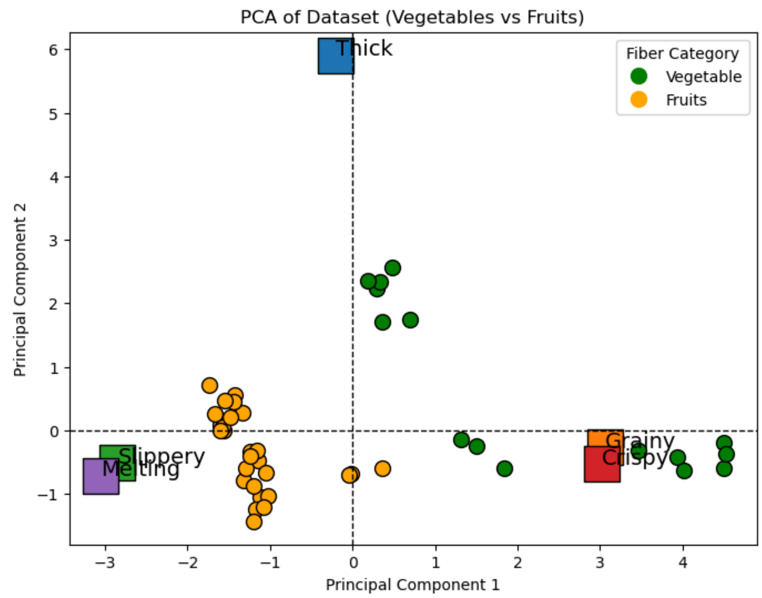


Figure 6. *PCA plot by dividing fiber groups into vegetable and fruit*

Corrplot was made to investigate the relationship between dependent variables. From which we could observe that there's a positive relation between melting and slipperiness, graininess and crispiness. Besides, negative relations between melting and grainy and crispy, slippery and grainy and crispy can be observed.

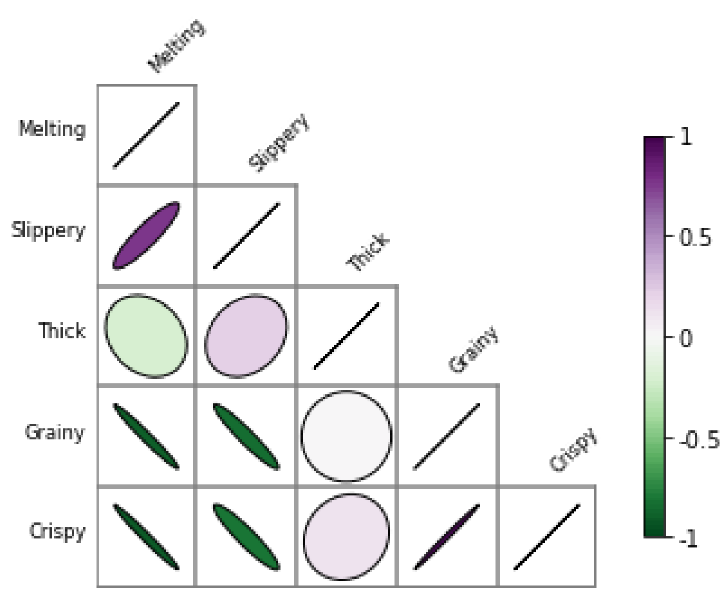


Figure 7. *Corrplot showing relations between dependent variables, where the more elliptical the pattern is, the stronger the relationship it is.*

To study the effect of treatment conditions on all the product characteristics, MANOVA was performed. With the result shown in table 2, it can be seen that all treatments have significant impact on the characteristics.

Table 2. MANOVA result

Treatment	P value	Significance
Concentration	0.002	**
Fiber type	0.000	**
Pressure	0.000	**

ns:P- value >0.05, "":P- value <=0.05 and "**":P- value <=0.01

Univariate analysis of the interested dependent variable

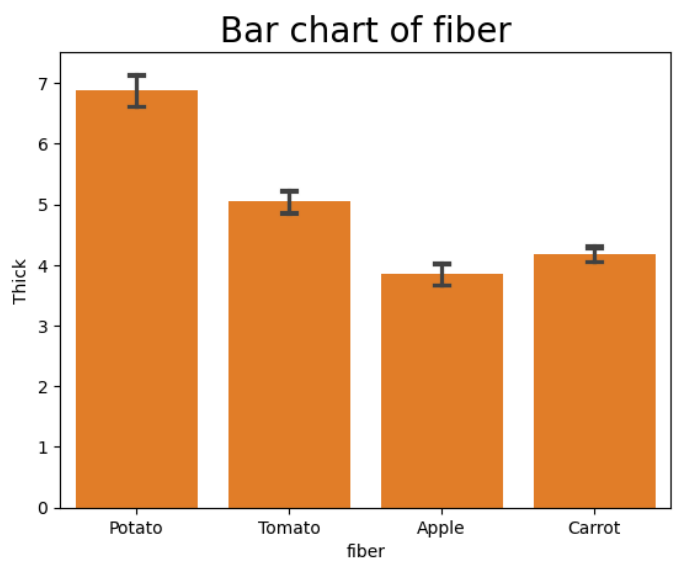


Figure 8. *Kruskal-Wallis bar chart for the thickness of four fiber types.*

To investigate the effect fiber type has on sauce characteristic, univariate analysis was performed. The significance was studied by Kruskal-Wallis test shown in table 3 and a barchart for thickness is plot. For the thickness we're interested in, a p value of $2.05*10^{-7}$ was present, indicating the fiber type has significant influence on this variable. Also from the bar chart, potato fiber products have higher average than all the other fiber types.

Table 3. Kruskal-Wallis statistic result for fiber type effect on characteristics

Characteristics	P value	Significance
Crispiness	$8.39*10^{-7}$	**
Thickness	$\mathbf{2.05*10^{-7}}$	**
Graininess	$2.32*10^{-6}$	**
Melting	$1.62*10^{-6}$	**
Slipperiness	$4.03*10^{-7}$	**

ns:P- value >0.05, “”:P- value <=0.05 and “**”:P- value <=0.01

To further investigate if thickness is infected by other independent variables, the same tests were done for pressure and concentration. The results are shown in table 4 and figure 9. From the bar chart of pressure, not very significant difference in average can be seen, which aligns with the result from Kruskal-Wallis test. As for concentration, slight differences with variance can be observed.

Table 4. Kruskal-Wallis statistic result for treatment condition on thickness

Condition	P value	Significance
Pressure	0.101	ns
Concentration	0.003	*

ns:P- value >0.05, “”:P- value <=0.05 and “**”:P- value <=0.01

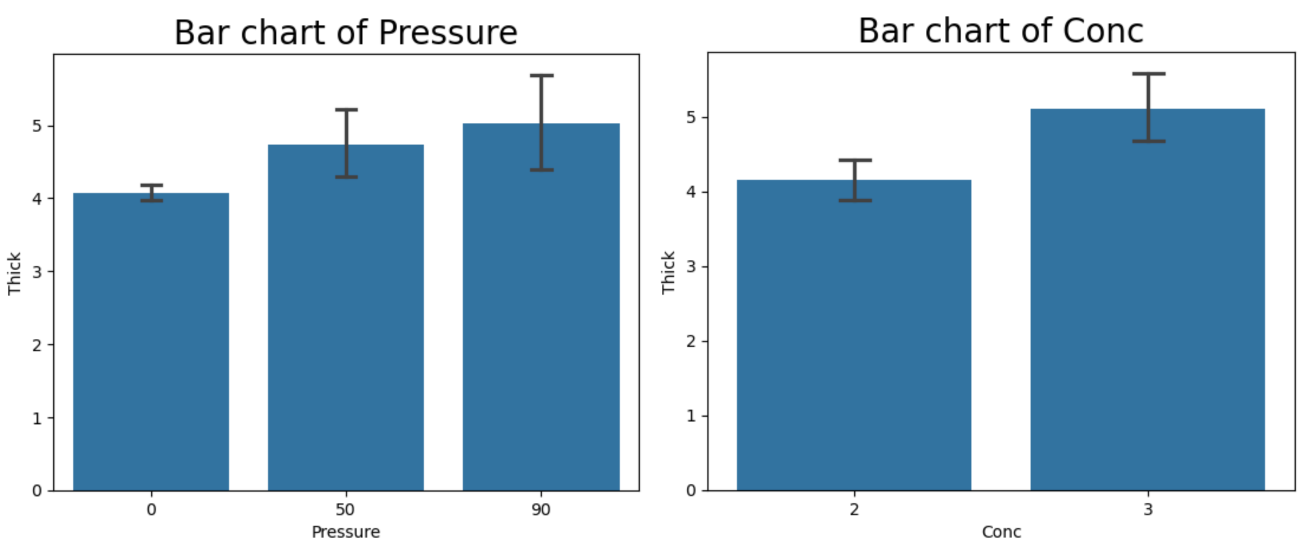


Figure 9. *Bar chart for thickness (left) and concentration (right)*

Discussion

The analysis of this study implies the significant impact of fiber type on the physical characteristics of the sauces, particularly in relation to thickness, which was the main focus of the hypothesis. The results from both PCA and Kruskal-Wallis tests offer clear evidence supporting the hypothesis that the different fiber types significantly affect product characteristics.

From the principal component analysis shows the relationship between sensory parameters and fiber types. It was observed that fibers cause great impact on the taste of the sauce. Potato based products are thicker, carrot based products are more crispy and grainy, while apple based are smoother and easier to melt.

Grouping the fiber types into fruits (tomato, apple) and vegetables (potato, carrot) and observing their clustering in the PCA is consistent with findings from a study by Wang et al. (2019), who examined the influence of fruit and vegetable fibers on the sensory characteristics of bakery products. They also found that vegetable fibers contribute to thicker textures, while fruit fibers lead to smoother, more fluidic products.

The Kruskal-Wallis test gives statistical support to these findings. The p-values obtained for the effect of fiber type on various sensory parameters were below 0.01, indicating it plays a critical role in affecting the sensory characteristics of sauces.

Interestingly, the pressure doesn't have as strong an impact on the characteristics of the sauces, as indicated by both the PCA (shown in figure A.1 in Appendix) and statistical tests. While concentration has only a slight impact. This indicates fiber type is the most principal factor in influencing sauce properties. This outcome aligns with our hypothesis and underscores the importance of ingredient composition (specifically fiber types) over external factors like pressure in determining sauce texture.

Since there's some correlation between sensory parameters themselves. It might be possible to predict one based on the other (if they're strongly correlated), and make it easier to focus on key parameters when developing. For example, if crispiness and graininess are linked, developers could focus on adjusting one parameter, knowing that it will likely affect the other in predictable ways.

Also, if dummy variables are used in the PCA plot, the result would be pretty different (shown in figure A.2 in Appendix). The reason might be that the addition of the columns are treated as individual features by PCA, and they can influence the variance of the data. Hence, it might not be considered as a good strategy to perform PCA in this situation.

Future study can involve studying the cause of differences in fruits fiber product and vegetable fiber product, and furthermore even the nutrient differences in those two types of fibers.

Conclusion

This study confirmed that fiber type influences the sensory and physical characteristics of organic sauces. Vegetable fibers (potato and carrot) contributed to thicker, grainier textures, while fruit fibers (tomato and apple) led to smoother, slipperier sauces. PCA and Kruskal-Wallis tests highlighted that fiber type, rather than pressure, plays a key role in defining sauce texture.

These findings provide valuable insights for optimizing sauce formulations based on fiber choice, supporting future product development in the organic and health-conscious food market.

Appendix

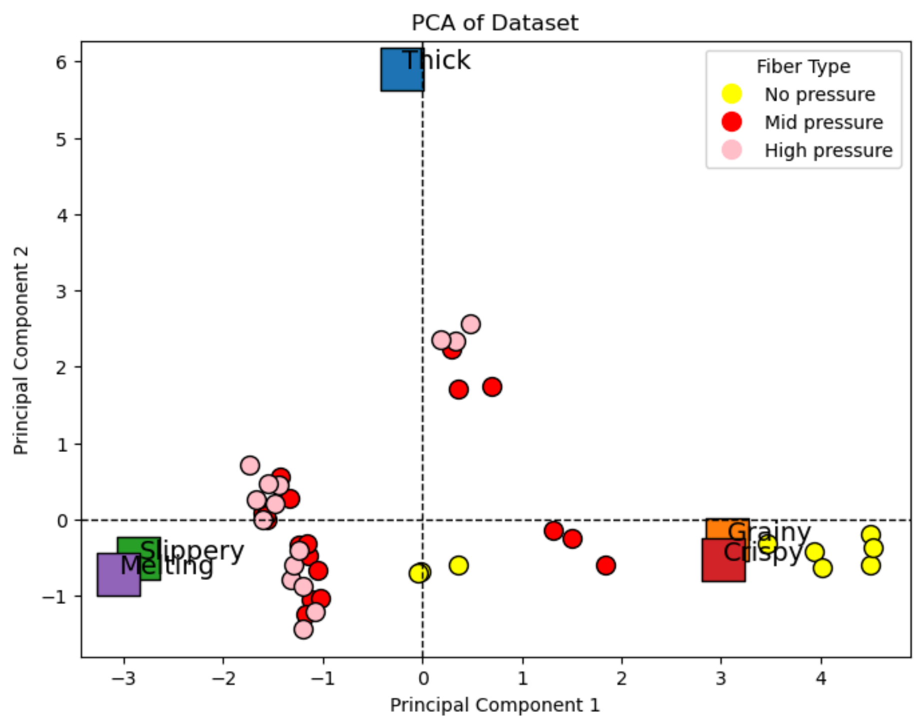


Figure A.1 *PCA analysis grouped by pressure treatment*

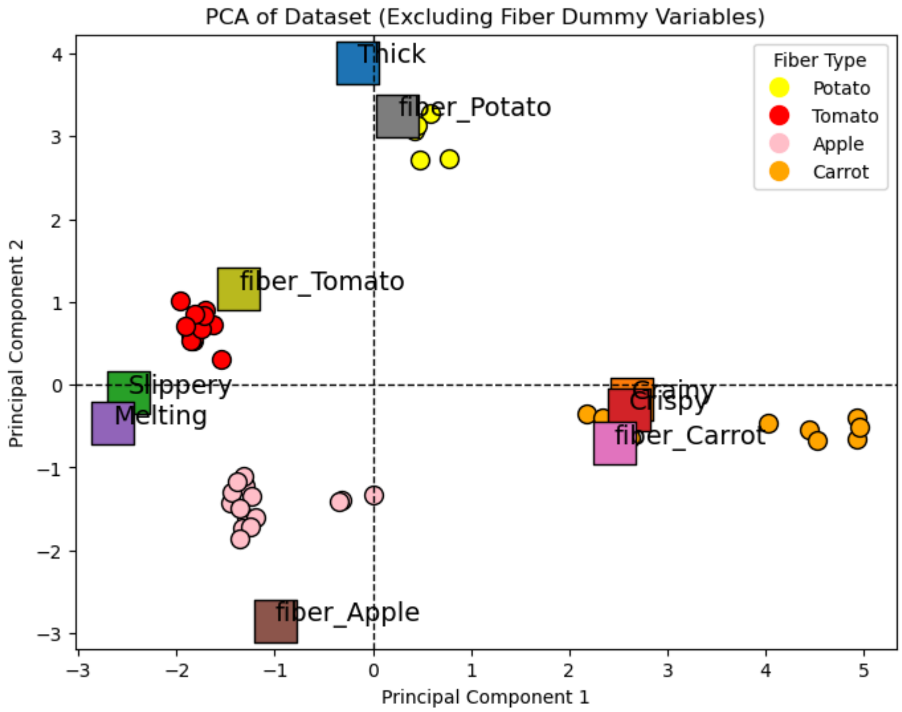


Figure A.2 *PCA analysis with dummy variables*

Acknowledgments

AI was used in code writing as well as report language improvement.

References

Wang, J., Rosell, C. M., & Benedito de Barber, C. (2019). "Effect of the addition of different dietary fibers on dough performance and bread quality." *Food Chemistry*, **79**(2), 221-226.